



City Research Online

City, University of London Institutional Repository

Citation: Batchelor, R. (2010). Worst-Case Scenarios in Forecasting: How Bad Can Things Get?. *Foresight: The International Journal of Applied Forecasting*, 18, pp. 27-32.

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/7678/>

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

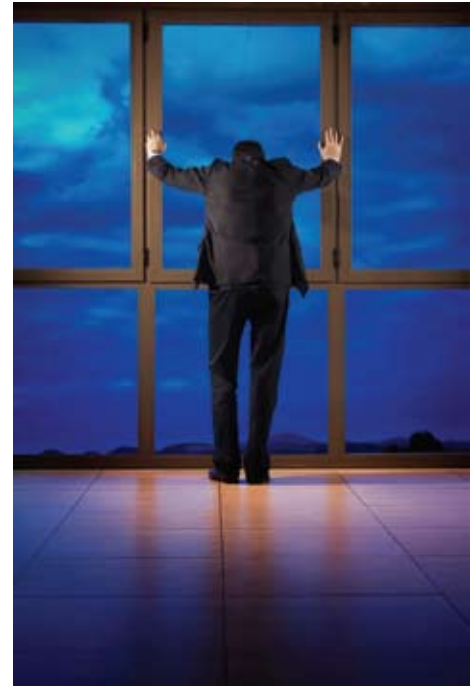
Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Worst-Case Scenarios in Forecasting: How Bad Can Things Get?

Roy Batchelor

Worst-Case Damage from a Blowout in the Gulf of Mexico

According to news reports, the main contingency plan foresaw a blowout with a worst-case spill of 40 million gallons in total. During the first three months, millions of gallons have been gushing per day. Regulatory agencies based their efforts on worst-case scenarios that weren't nearly worst case.



PREVIEW. Roy Batchelor, Foresight's Financial Forecasting Editor, explains that conventional business-forecasting models are not set up to tell us about the impacts of extreme events—hence, their worst-case forecasts are liable to be less severe than the worst that plays out in the future. While recognition of this caveat is important in itself, there are ways to model the impacts of extreme events and thus derive a realistic indication of downside risk.

INTRODUCTION

Throughout the recent recession, companies have watched sales and production shrink by unprecedented amounts. Month after month, outturns have not merely fallen short of central forecasts but have crashed through the lower limits of prediction intervals churned out by statistical models. Conventional business-forecasting systems are just not set up to tell us about extreme events.

Downside risk has long been a central concern of financial forecasters. In the financial markets, high volatility in prices means large potential losses for investors, and risk-averse hedgers will pay more for insurance against adverse events. Option pricing theory ties the cost of insurance directly to the forecast of volatility of future

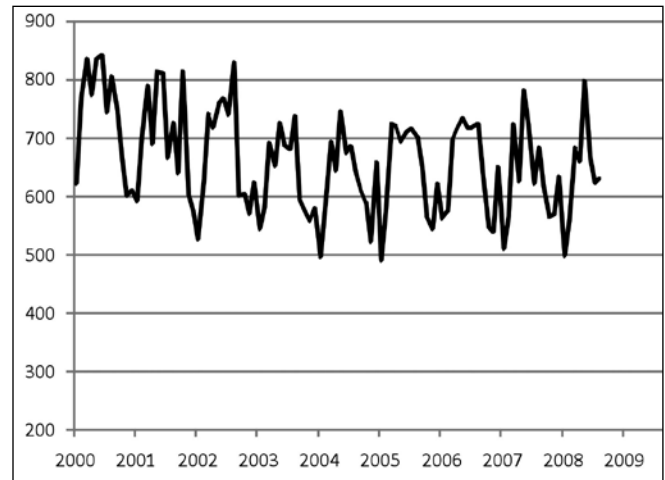
price changes. A massive academic and practitioner literature has sprung up, focused on getting good predictions of whether share prices, currencies, and commodities are likely to become more or less volatile in the future.

Volatility changes can be forecast. Day-to-day price changes are close to random, but the volatility of these price changes is *serially correlated*: If there is a big price change (up or down) on one day, it is more likely than not that there will be a big price change the following day (down or up, we don't know which).

More recently, downside risk has become an important issue for business forecasters who are concerned with future sales, rather than prices on the financial markets. Most

companies try to define “worst-case scenarios” for sales. Models of inventory control rely heavily on estimates of the future volatility of demand. Yet somehow, the well-developed technology of volatility forecasting has not been transferred into the business domain. To see what that would involve, I have set out below the procedure for defining a worst-case scenario that would follow from reading a basic business forecasting text. Then I illustrate how the transfer of a small piece of volatility forecasting technology – the GARCH variance model – from a financial to a business-forecasting environment can help quantify downside risk.

Figure 1. New Car Sales, January 2000- August 2008 (thousands)



A STANDARD VIEW OF NEW CAR SALES

Consider the point of view of a forecaster trying to predict new-car sales in the U.S. through the 2008-9 recession. Imagine we are in early September 2008. We have a preliminary estimate of August sales of around 630 thousand vehicles. This was about 10% lower than August of a year earlier, but still well within the range of 500-800 thousand that had been the norm since 2000. Visual inspection of the monthly data on Figure 1 suggests that a worst-case scenario would be monthly sales below 500 thousand, an event that happened in only three of the previous 103 months.

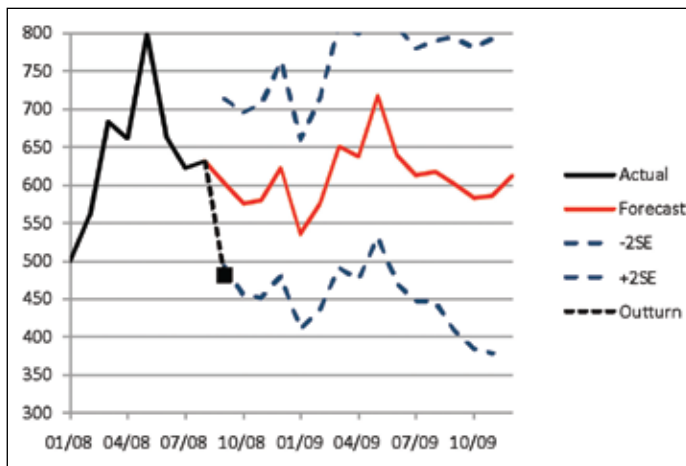
To generate a forecast for September 2008 onwards, we need a model for car sales. I have used a conventional time-series representation using data back to 1980, a seasonal ARIMA model. The forecasts from this model are shown on Figure 2, in the form that is generated by most standard business-software packages. This shows expected sales (red line) and upper and lower bounds to the 95% prediction interval. The idea is that, in 95% of forecast months, sales should lie within these limits. Sales are forecast to be around 600 thousand cars per month, with some seasonal fluctuations. The bad case is for sales to fall below 494 thousand in September. The estimated reliability of the central forecast is reflected in its standard

Key Points

- Standard statistical models provide misleading evidence for defining the “worst-case” possibilities when there is a serious shock to the business, as happened to the world’s Wall and Main streets during this past recession.
- Serious shocks normally increase the volatility of sales; this extra volatility has to be modeled appropriately. When we are looking at extreme events, it is important to understand that the worst event that has already happened in the sample is not as bad as the worst event that can possibly happen.
- One modeling approach used in finance is called GARCH, a method that assumes that volatility and hence expected forecast errors increase after a large external shock.
- Through a case study of automobile sales, I illustrate how a GARCH model provides a more realistic forecast of the downside risk facing this industry.

error, which we could loosely define as the size of a typical monthly error. Larger values of the standard deviation imply less reliability. This lower bound is around two standard errors below the expected level of sales, so the standard error of our one-month-ahead forecast is about $1/2 \times (600 - 494) = 53$ thousand cars, or about 9% of the central forecast.

Figure 2. Forecasts and Outturn for September 2008



What actually happened in September 2008, the month of the Lehman bankruptcy, was that consumer confidence plummeted, spending contracted across the economy, and car sales fell to an all-time low of 481 thousand, below the model's bad-case estimate. Of course, we expect there to be two to three months in every decade when sales fall below the lower prediction bound. This was an extreme adverse event, and leads to our key question: Given such a shock, how much worse can things get?

Let's stick with our conventional model and make a new set of forecasts for October 2008 and beyond, in light of information about the collapse in sales in September. The new path of expected sales and lower bound to the 95% prediction interval are shown in Figure 3. Sales are predicted to rebound to 505 thousand and then continue along a path somewhat lower than had been forecast a month earlier. Very sensibly, part of the fall in sales in September is treated as a temporary effect that will be offset in the following month, and

part is treated as a signal that the underlying level of the series has permanently fallen.

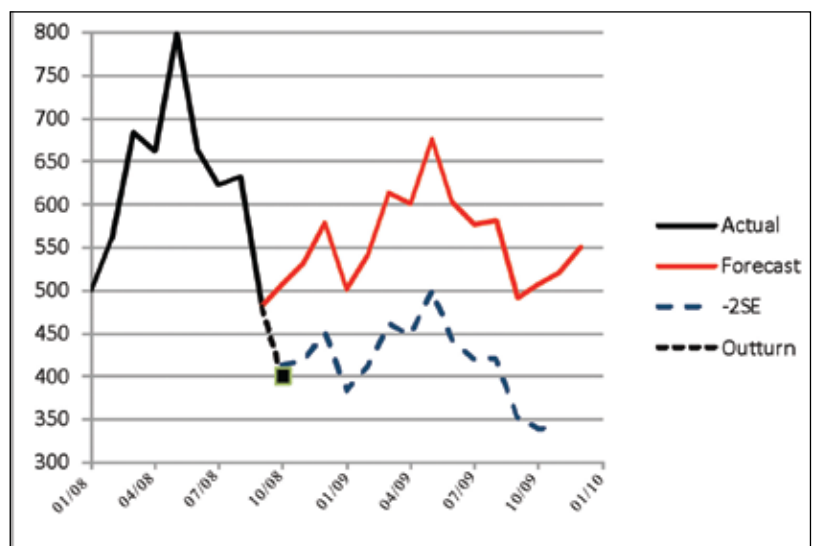
Less sensible is what happens to the prediction interval. The bad case is now for sales to be 413 thousand. This is 92 thousand below the mean forecast, suggesting a standard error of $92/2 = 46$ thousand. This is actually lower than a month earlier, and still only 9%

of the new mean forecast. So, after the worst shock to car sales in living memory, the standard business-forecasting model suggests that our forecasts will be just as reliable as they were before the shock happened!

To underline how unrealistic this is, Figure 3 also shows that sales in October again fell below the model-generated lower prediction bound. Two successive outcomes below the lower bound is a very unusual occurrence and should make

us reconsider how our bad-case forecast has been constructed. We could keep going – but, month after month, the standard error of the one-month-ahead forecast would stay stuck at about 9% of the forecast level, regardless of whether the economic conditions were calm or stormy.

Figure 3. Forecasts and Outturn for October 2008



Box 1. The GARCH Model

Suppose we are using time-series data to forecast a target variable y_t , based on a set of predictors x_t . Our standard regression model is $y_t = bx_t + u_t$ where b is a vector of coefficients, and the u_t are regression residuals (“shocks” that cannot be explained by the predictors x). The standard assumption is that the u_t are normally distributed and have a standard deviation σ that is constant over time.

The GARCH model allows σ to vary over time. Specifically, if at time $t-1$ there was a big shock (so the squared residual u_{t-1}^2 is large), then volatility σ will rise. Conversely, if the last residual was small, then σ will fall. The exact formula used is $\sigma_t^2 = a_0 + a_1 u_{t-1}^2 + a_2 \sigma_{t-1}^2$. The larger the size of a_1 relative to a_2 , then the greater the influence will be of the latest shock on our new estimate of volatility. The closer a_2 is to 1, the more long-lived will be the effect of a large shock on the volatility of y in subsequent time periods.

The distribution of the shocks u_t need not be normal, and in the case of car sales follows a t-distribution with 8 degrees of freedom, showing that there are many more extreme events. In this case, the 95% prediction interval is not $\pm 2\sigma$, but $\pm 2.3\sigma$, making the margin of uncertainty wider and the worst-case scenario even worse.

For the U.S. car market, a_1 is 0.20 and highly significant, meaning that an unexpected 10% fall in sales leads to a rise of $\sqrt{(0.2 \cdot 10^2)} = 4.5\%$ in the standard deviation of car sales in the following month. Therefore, if volatility had been around 8% of sales, an unexpected 10% fall in sales would cause volatility to rise to 12.5% of car sales – exactly what happened between September and October 2008.

The coefficient a_2 is 0.6, so the shock has a half-life of $1/(1-0.6) = 2.5$ months. That is, after the initial large impact of the shock, in the absence of further shocks volatility will die away towards a baseline level quite quickly over the following months.

There are many variants of the GARCH model. With car sales, reactions to good news and bad news are the same. This is why volatility rose after the unexpected surge in sales in August 2009. However, with stock prices, a large fall in the market increases volatility much more than a large rise, so when modeling the stock market the coefficient a_1 would be higher for negative shocks than for positive shocks.

A relatively nontechnical review of GARCH is given in Robert Engle (2001), GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics, *Journal of Economic Perspectives*, 15, 4, 157-168, downloadable at <http://pages.stern.nyu.edu/~rengle/Garch101.doc>

TIME-VARYING VOLATILITY IN CAR SALES

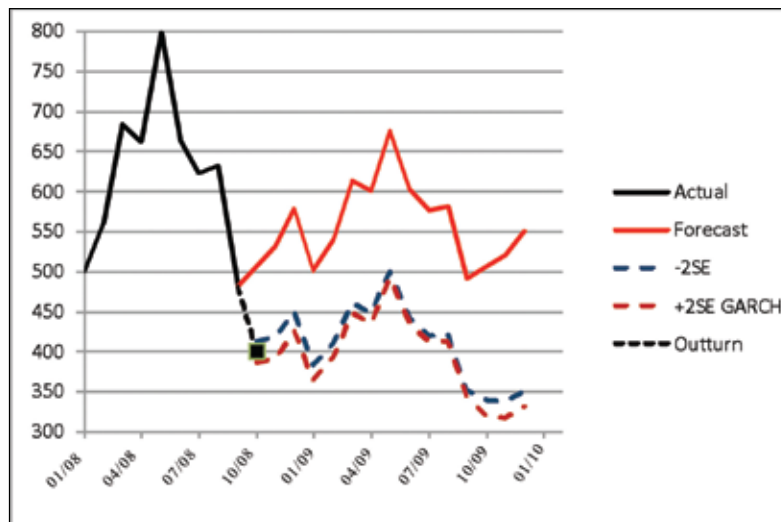
The problem is that conventional time-series and regression models assume that the distribution of shocks to the system stays *unchanged over time*. Sometimes shocks are large and sometimes small, but these are treated as random draws from an underlying probability distribution that has a constant volatility. Prediction intervals from these models are based on an estimate of the average volatility of residuals over the whole sample used in model estimation. When a new observation appears, no matter how extreme it is, it has only a marginal effect on the estimated variance of residuals, and hence scarcely affects prediction intervals.

This assumption of constant volatility of shocks – called *homoscedasticity* – is not one that would be entertained in any model of financial markets, where it is plain that periods of steady growth in markets are punctuated by booms and crashes, during which volatility rises sharply. Since the Nobel Prize-winning economist Rob Engle developed the so-called ARCH (*autoregressive conditional heteroscedasticity*) model in the 1980s, it has become standard to characterize financial time series by some variant of this model. Financial economists are also reluctant to assume that shocks are normally distributed, preferring “fat-tailed” distributions such as the t-distribution, which allow extreme events to occur more frequently than suggested by the normal curve.

In the most popular “generalized ARCH” or GARCH model, there is some long-term underlying average variance, but in the short term the variance of potential shocks can rise above this underlying level if there is an unexpectedly large shock (= large forecast error) to the series being modeled. I describe the GARCH model in Box 1.

Regression packages churn out tests for normality in residuals (Jarque Bera test) and for ARCH errors (Engle’s Lagrange multiplier test). These are often ignored

Figure 4. Lower 95% Prediction Bounds: Conventional v. GARCH



because they do not bias central forecasts. They are critical, however, for constructing prediction intervals. Therefore, we can easily test whether shocks to car sales can be described by a GARCH model and whether their distribution contains more extreme events than normal. The answer is yes, and yes. Volatility in car sales does rise after unusually large and unexpected increases and decreases in sales. When shocks occur, there are more large changes than the normal curve would lead us to expect.

Let's revisit the forecasts for September and October 2008, using the same ARIMA model for car sales but allowing the distribution of shocks to be fat-tailed and have time-varying volatility. Figure 4 shows the effect of the GARCH assumption on the mean forecast and the lower prediction interval bound. The mean forecast is unchanged. However, the large shock that occurred in September 2008 has caused the model to revise sharply upwards its estimate of the likely volatility of future shocks. The lower prediction bound for October is now 386 thousand, much lower than the 413 thousand estimated by the conventional model.

The outturn of 400 thousand is now inside rather than outside the prediction

interval, and in that sense is less surprising. The fact that the October outturn is well

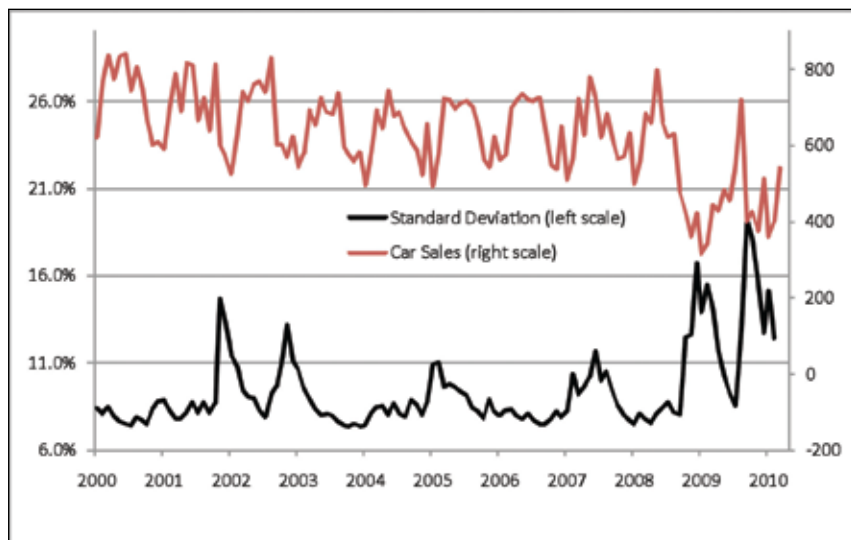
below its expected value also means that, when we make our forecast for November, the GARCH model will again predict that volatility will be high, the prediction interval large, and the bad case will again be pretty bad.

Figure 5 shows the GARCH estimates of how the standard deviation of

shocks in the car market has changed from month to month since 2000. Although the average volatility is indeed around 9%, it can change drastically from year to year, and the very steep rise to over 15% in the recent recession shows that models of car sales that neglect changes in volatility provide us with a very poor guide to the risks faced by producers and dealers.

Note, by the way, that the most recent peak in volatility was due not to a collapse in sales, but to the splurge of buying in August 2009 in response to the "cash for clunkers"

Figure 5. GARCH estimates of time-varying volatility (standard deviation) of new-car sales



scheme that subsidized the replacement of old cars by new, more fuel-efficient vehicles.

CONCLUDING REMARKS

I have looked here at just one way of refining estimates of prediction intervals, using a simple model of time-varying volatility. There are many other ingenious devices for looking at downside risk in the financial-risk manager's toolkit. For example, some analysts ignore all but the most extreme events and use special extreme value distributions to approximate the shape of the left tail of probability distributions. These methods are viable only if we have many observations on extreme events, and this in turn depends

on the availability of high frequency, daily, or intraday data over a long time period – conditions that rule out most mainstream business-forecasting applications.

Business forecasters don't often take prediction intervals seriously and have some incentives to keep quiet about them. Most of the time, the intervals look very scary. With moderate sample sizes, the conventional 95% prediction interval for three to four steps ahead typically encompasses the whole range of historic data, and colleagues and clients might be tempted to conclude that you are saying, "Anything can happen."

Unfortunately, this is true. Indeed, results from forecasting competitions consistently tell us that, if anything, statistical prediction intervals are too narrow, since we can rarely identify the true model driving our data, and all series are subject to unforecastable structural change.

Frank Sinatra had the lyric "The best is yet to come" inscribed on his tombstone. When forecasting extreme events, it is important to understand, as well, that the worst is yet to come. The worst thing that has already happened in the sample is an upper estimate of the worst thing that can possibly happen. The honest answer to the question "How bad can things get?" is that they can always be worse than they have ever been before.

The value of models with time-varying volatility is that they help us quantify *exactly how much worse* things can get: whether right now we are confronted with a low or a normal degree of risk, or whether – as in September 2008 – volatility is unusually high, and one of these unprecedentedly bad outcomes is most likely to occur.

When forecasting extreme events, it is important to understand, as well, that the worst is yet to come. The worst thing that has already happened in the sample is an upper estimate of the worst thing that can possibly happen. The honest answer to the question "How bad can things get?" is that they can always be worse than they have ever been before.



Roy Batchelor is *Foresight's* Financial Forecasting Editor. His day job is professor of banking at Cass Business School, City University of London. Roy consults extensively for public and private organizations and has spent considerable time

recently helping Dubai overcome its financial difficulties.

r.a.batchelor@city.ac.uk

Copyright of Foresight: The International Journal of Applied Forecasting is the property of International Institute of Forecasters and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.